Journal of Zhejiang University-SCIENCE B (Biomedicine & Biotechnology) *2023 24(10):883-895*
www.jzus.zju.edu.cn; www.springer.com/journal/11585
E-mail: jzus_b@zju.edu.cn

Research Article
https://doi.org/10.1631/jzus.B2200634

Molecular dynamics simulation reveals DNA-specific recognition mechanism via c-Myb in pseudo-palindromic consensus of *mim-1* promoter

Jinru WENG[1*], Shuo YANG[2*], Jinkang SHEN[1], Hongsen LIU[1], Yuzi XU[1], Dongyun HAO[3✉], Shan WANG[1✉]

[1]*Stomatology Hospital, School of Stomatology, Zhejiang University School of Medicine, Zhejiang Provincial Clinical Research Center for Oral Diseases, Key Laboratory of Oral Biomedical Research of Zhejiang Province, Cancer Center of Zhejiang University, Engineering Research Center of Oral Biomaterials and Devices of Zhejiang Province, Hangzhou 310000, China*
[2]*Key Laboratory for Molecular Enzymology and Engineering of the Ministry of Education, Jilin University, Changchun 130021, China*
[3]*Institute of Agricultural Biotechnology, Jilin Academy of Agricultural Sciences (JAAS), Changchun 130033, China*

Abstract: This study aims to gain insight into the DNA-specific recognition mechanism of c-Myb transcription factor during the regulation of cell early differentiation and proliferation. Therefore, we chose the chicken myeloid gene, mitochondrial import protein 1 (*mim-1*), as a target to study the binding specificity between potential dual-Myb-binding sites. The c-Myb-binding site in *mim-1* is a pseudo-palindromic sequence AACGGTT, which contains two AACNG consensuses. Simulation studies in different biological scenarios revealed that c-Myb binding with *mim-1* in the forward strand (complex F) is more stable than that in the reverse strand (complex R). The principal component analysis (PCA) dynamics trajectory analyses suggested an opening motion of the recognition helices of R2 and R3 (R2R3), resulting in the dissociation of DNA from c-Myb in complex R at 330 K, triggered by the reduced electrostatic potential on the surface of R2R3. Furthermore, the DNA confirmation and hydrogen-bond interaction analyses indicated that the major groove width of DNA increased in complex R, which affected on the hydrogen-bond formation ability between R2R3 and DNA, and directly resulted in the dissociation of DNA from R2R3. The steered molecular dynamics (SMD) simulation studies also suggested that the electrostatic potential, major groove width, and hydrogen bonds made major contribution to the DNA-specific recognition. In vitro trials confirmed the simulation results that c-Myb specifically bound to *mim-1* in the forward strand. This study indicates that the three-dimensional (3D) structure features play an important role in the DNA-specific recognition mechanism by c-Myb besides the AACNG consensuses, which is beneficial to understanding the cell early differentiation and proliferation regulated by c-Myb, as well as the prediction of novel c-Myb-binding motifs in tumorigenesis.

Key words: c-Myb; DNA-specific recognition mechanism; Molecular dynamics simulation; DNA major groove width; Electrostatic potential

1 Introduction

As a transcription factor, c-Myb plays an essential role in regulating cell development (Oh and Reddy, 1999), including the processes of stem cell differentiation and proliferation (Ramsay and Gonda, 2008), hematopoiesis (Greig et al., 2010), airway epithelial cell differentiation (Pan et al., 2014; Pardo-Saganta et al., 2015), bone formation (Zhao et al., 2011; Bhattarai et al., 2013; Tan et al., 2013; Lee et al., 2016), proliferation of neural progenitor cells in the regions of brain and colonic crypts (Malaterre et al., 2008; Greig et al., 2010), and promotion of T cell stemness (Maurice et al., 2007; Gautam et al., 2019; Heuser and Gattinoni, 2022). Myb transcription factors belong to a proto-oncogene product super family that has been identified in all biological kingdoms (Katzen et al., 1985; Nishina et al., 1989; Rosson and Reddy, 1986; Paz-Ares et al., 1987). As a proto-oncogene, overexpression or rearrangement of c-Myb has been

✉ Shan WANG, shan_wang@zju.edu.cn
Dongyun HAO, dyhao@cjaas.com
* The two authors contributed equally to this work
Shan WANG, https://orcid.org/0000-0002-9035-2290
Dongyun HAO, https://orcid.org/0000-0001-8550-0992

Received Dec. 7, 2022; Revision accepted Feb. 7, 2023;
Crosschecked Sept. 7, 2023; Published online Sept. 23, 2023

detected in a wide range of human cancers, including leukemia (Mansour et al., 2014; Clesham et al., 2022), breast cancer (Gonda et al., 2008), colon cancer (Torelli et al., 1987; Abaza et al., 2004; Hugo et al., 2006), pancreatic cancer (Wallrapp et al., 1997), head and neck cancer (Mitani et al., 2010; Lu et al., 2013; Ni et al., 2015; Xu et al., 2019), osteosarcoma (Aggarwal and Bhavesh, 2021), and other solid tumors (Patt et al., 1993; Miyazaki et al., 2012; Lu et al., 2013; Cicirò and Sala, 2021). Mansour et al. (2014) showed that an insertion in the precise noncoding site plays a pivotal role in oncogenesis in T-cell acute lymphoblastic leukemia, which results in the formation of a novel binding site of c-Myb and creates a super-enhancer upstream of T-cell acute leukemia protein 1 (*TAL1*) oncogene. Fuglerud et al. (2017) demonstrated that D152V, one amino acid mutant in the DNA-binding domain (DBD) of c-Myb, can impair the ability of c-Myb to regulate hematopoietic cell differentiation, which affects histone binding and increases chromatin accessibility. This raises the question of what might be the DNA-specific-binding mechanism of c-Myb transcription factor.

In general, DNA-specific recognition mechanisms by a transcription factor are classified into two types, one involving the formation of hydrogen bonds with specific bases, mainly in the major groove, and the other entailing the sequence-dependent deformations of the DNA helix (Rohs et al., 2009). The DBD of c-Myb consists of three tandem repeats (R1, R2, and R3), each containing three α-helices (Biedenkapp et al., 1988). The second and third helices of each repeat form a helix-turn-helix (HTH) structure with three regularly spaced tryptophan (or hydrophobic) residues that are involved in the Myb–DNA interaction (Ogata et al., 1996; Xu et al., 2019; Kawasaki and Oda, 2021). Meanwhile, the third helices of R2 and R3 (R2R3) are regarded as the "recognition helix," since they directly and cooperatively bind to a conserved base sequence, AACNG (Ogata et al., 1992, 1994, 1995, 1996; Kanei-Ishii et al., 1994). The three residues in the recognition helices of R2R3 play important roles in the DNA-binding mechanism, which are Asn183 (R3), Lys182 (R3), and Lys128 (R2) (Ogata et al., 1994).

Myb transcription factors are characterized by two kinds of target-binding sequences: (1) pseudo-palindromic sequence with dual-Myb-binding sites (e.g., mitochondrial import protein 1 (*mim-1*), MYC, and T cell receptor (TCR)), and (2) only one consensus sequence with a single Myb-binding site (e.g., Moebius syndrome 1 (MBS1), target of Myb1 membrane trafficking protein (TOM1), and cluster of differentiation 4 (CD4)) (Ness, 1996; Wolff, 1996). As one of well-studied Myb target genes, the chicken myeloid gene *mim-1* is specifically activated by c-Myb in myelomonocytic cells (Ness et al., 1989; Graf, 1992). The c-Myb-binding site in *mim-1* is a pseudo-palindromic sequence AACGGTT (Ness et al., 1989), which contains two potential AACNG consensuses in the opposite orientations (5'-AACGG-3' in the forward strand and 5'-AACCG-3' in the reverse strand; Fig. 1a).

In this study, we chose the *mim-1* as a representative of the pseudo-palindromic consensus to investigate the specific recognition mechanism of c-Myb. Six independent molecular dynamics (MD) simulations were performed on three complexes of c-Myb binding with *mim-1* in the forward strand (complex F), c-Myb binding with *mim-1* in the reverse strand (complex R), and c-Myb binding with MBS1 in the

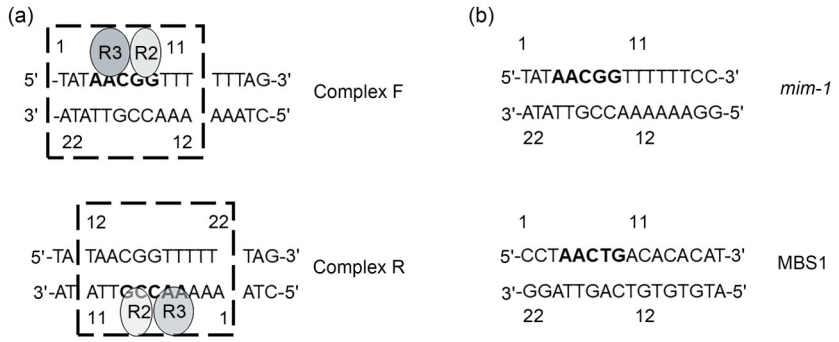

Fig. 1 Construction schemes of *mim-1* complexes for MD simulations. (a) Construction scheme of c-Myb and *mim-1* complexes in MD simulations; (b) DNA-binding fragments (16 bp) by *mim-1* and MBS1. MD: molecular dynamics; *mim-1*: mitochondrial import protein 1; MBS1: Moebius syndrome 1.

forward strand at both 300 and 330 K (Table S1). The principal component analysis (PCA) dynamics trajectory, electrostatic potential, DNA confirmation, and hydrogen-bond interaction analyses were performed on these six simulation systems. Interestingly, our analyses indicated that the electrostatic potential, major groove width, and hydrogen bonds made major contribution in the DNA-specific recognition mechanism by c-Myb, which was further confirmed by steered molecular dynamics (SMD) simulation studies on complexes F and R. Finally, we performed kinetic pattern analyses of c-Myb binding to *mim-1* and MBS1 in vitro to determine the binding specificity of c-Myb with *mim-1*. This study provides new insights into understanding the DNA-specific recognition mechanism by c-Myb.

2 Materials and methods

2.1 Preparation of structure

The refined average solution structure of c-Myb binding to DNA of MBS1 (Fig. 1b) was obtained from the Protein Data Bank (PDB code: No. 1MSE) (Bernstein et al., 1978). The *mim-1* DNA structures of 5'-TAACGG-3' and 5'-AAACCG-3' (Fig. 1a) were created using Open Babel (O'Boyle et al., 2011), and docked into the active site of c-Myb (PDB code: No. 1MSE) using AutoDock (Morris et al., 2009), deriving two c-Myb and *mim-1* complexes that were named complex F and complex R. The binding-site positions were shown in Fig. 1b for complexes F and R.

2.2 Molecular dynamics simulations

MD simulations were performed using the GROMACS program (version 2022.2) (Berendsen et al., 1995; Hess et al., 1997), using the modified AMBER99SB-ILDN force field (Lindorff-Larsen et al., 2010). The six independent simulation systems were established at two different temperatures of 300 and 330 K for the 400-ns simulation time scale (Table S1). In addition, the system of R2R3 binding with MBS1 was established at 300 K. Each system was solvated in a dodecahedron periodic box containing the SPC216 water model (Teleman et al., 1987), which was extended to at least 10 Å (1 Å=1×10^{-10} m) between the model and the edge of water box. To neutralize the charge of water box, adequate numbers of Na^+ ions were added into each system (Table S1). The structure of each system was relaxed through 50 000 steps of energy minimization. Subsequently, the equilibration of each system was performed in two phases: 100-ps constant NVT (number of particles, volume, and temperature) simulation and 100-ps constant NPT (number of particles, pressure, and temperature) simulation. Afterwards, a 400-ns MD simulation was performed on each system. During MD simulation, the linear constraint solver (LINCS) algorithm (Hess et al., 1997) was used to constrain the lengths of all the bonds. The time step for the simulation was 0.002 ps. The thermostat and barostat coupling on each ensemble were respectively performed using the Nose-Hoover and Parrinello-Rahman methods. Long range electrostatic forces were treated using the particle-mesh Ewald (PME) method (Essmann et al., 1995). van der Waals forces were treated with a 1.0-nm cut-off value.

SMD simulations were set up for complexes F and R (Table S2) and carried out using the GROMACS program (version 2022.2) with the AMBER99SB-ILDN force field. Complexes F and R were solvated in a 6.6 nm×4.4 nm×12 nm cuboid box containing the SPC216 water model. To neutralize the charge of the water box, adequate numbers of Na^+ ions were added to each system (Table S2). The energy minimization and NPT equilibration were then performed for each system in the same way as previously described (Wang et al., 2011). The optimal biasing potential along Z was a harmonic potential with a force constant of 2000 kcal/(mol·nm). Afterwards, a 300-ps SMD simulation was performed on each system.

The analyses of trajectories were performed using the packages embedded in GROMACS and in-house Perl scripts. For the protein structure in each system, PCA (Lauria et al., 2009; Yang et al., 2009) was performed using the packages embedded in GROMACS. The detailed method was described in our previous work (Wang et al., 2011). The conformation of the DNA helix in each system was analyzed using Curves+ (Lavery et al., 2009). The interaction of each system was analyzed using the PDBePISA service (Krissinel and Henrick, 2007). The electrostatic potential calculations were analyzed using the PyMOL Molecular Graphics System 2.0 software (http://www.pymol.org/pymol).

2.3 Preparation of proteins and oligonucleotides

The c-Myb minimum DBD R2R3 (Leu90-Val193) was overexpressed in *Escherichia coli* BL21 (DE3) as described (Nishina et al., 1989; Tanikawa et al., 1993). The purified R2R3 was suspended in 100 mmol/L potassium phosphate buffer (pH 7.5) containing 20 mmol/L KCl, 0.1 mmol/L ethylene diamine tetraacetic acid (EDTA), 500 μg/mL bovine serum albumin (BSA), and 10 mmol/L dithiothreitol. The purity of R2R3 protein was inspected to be no less than 95% on sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE). A 16-bp pseudo-palindromic consensus-containing *mim-1* fragment (5'-TATAACG GTTTTTTAG-3') from the *mim-1* gene and the 16-bp MBS1 fragment (5'-CCTAACTGACACACAT-3') from the simian virus 40 enhancer were prepared by synthesizing both strands (Ness et al., 1989; Ording et al., 1994). The sequences and position numbering of *mim-1* and MBS1 were presented in Fig. 1b. After purification, the fragments were suspended in 1× sodium chloride-Tris-EDTA (STE) buffer (0.1 mol/L NaCl, 10 mmol/L Tris-HCl, and 1 mmol/L EDTA; pH 8.0), and annealed with the complementary strands. The resulting double-strand oligonucleotides were end-labeled with [γ-^{32}P] adenosine triphosphate (ATP) (Amersham Pharmacia Biotech, Piscataway, NY, USA) using T_4 polynucleotide kinase (TOYOBO, Osaka, Japan) and purified through a Sephadex G-50 column. The single-base substitution-mutant sequences of *mim-1* were prepared in the same way.

2.4 DNA-binding assay

The fluctuation of binding free energy by reason of the base-pair substitution was evaluated as:

$$\Delta\Delta G=\Delta G_{\text{mutant}}-\Delta G_{\text{wild-type}}=RT\ln(K_{\text{D, mutant}}/K_{\text{D, wild-type}}), \quad (1)$$

where the equilibrium dissociation constant K_D of R2R3 binding to DNA of *mim-1* was measured using the quantitative electrophoretic mobility shift assay (EMSA) as described previously (Hao et al., 1998, 2002). R is the gas constant with a value of 8.314 J/(K·mol) and T is the temperature of the reaction. K_D for the dissociation of protein from DNA was determined by binding titration analysis. The positive value of $\Delta\Delta G$ means a reducing binding affinity. The 10-fold reducing binding affinity corresponds to 1.3 kcal/mol of $\Delta\Delta G$ at room temperature.

3 Results and discussion

3.1 Comparison of conformational stability between complexes F and R

In order to understand the c-Myb specific recognition mechanism on the dual-Myb-binding sites, two complexes were generated, namely, complex F and complex R (Fig. 1a). The root mean square deviation (RMSD) of α-carbon (Cα) atoms of R2R3 as a function of simulation time was calculated for each system to assess the conformational stability of the protein during MD simulation. As shown in Fig. 2, three systems of R2R3 binding with MBS1, complex F, and complex R were reproducible at 300 and 330 K. All of the six complexes reached plateau in a 400-ns time scale (Fig. 2a, Table S3), except complex R at 330 K. The RMSD of complex R at 330 K raised drastically after 200 ns (Fig. 2b), which indicated that the conformation of complex R at 330 K underwent a drastic fluctuation during that time. To compare the interaction of R2R3 with DNA in complexes F and R around 200 ns at 330 K, we extracted three time-point conformations of them for observation, which were at 190, 200, and 400 ns (Fig. 3). At 190 ns, the DNAs in both complexes F and R were still interacting with R2R3 (Figs. 3a and 3d). Interestingly, after 200 ns, the interaction of DNA and R2R3 in complex F was still kept (Figs. 3b and 3c), while the DNA in complex R had dissociated from R2R3 (Figs. 3e and 3f). This indicated that the interaction of R2R3 with DNA in the reverse strand was instable, as compared to the interaction of R2R3 with DNA in the forward strand.

In order to investigate the conformational variation of R2R3 during the whole simulation process, we analyzed the gyration radiuses of complexes F and R at 300 and 330 K as a function of the 400-ns simulation time. As shown in Fig. 4, there was an obvious gap in the gyration radiuses between complexes F and R during 160 to 260 ns at 330 K, and the average gyration radiuses of complexes F and R were 1.65 and 1.54 Å, respectively. Thus, we speculated that the R2R3 of complex R underwent a conformational variation in this period, and then recovered to the original conformation. Next, the root mean square fluctuations (RMSFs) of Cα per residue were analyzed for complexes F and R at 300 and 330 K. The two RMSFs between complexes F and R almost overlapped at

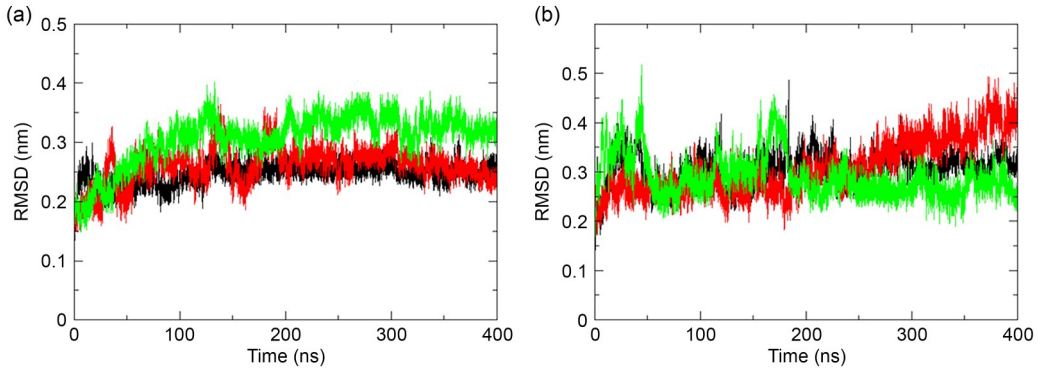

Fig. 2 Root mean square deviation (RMSD) profiles. (a) The α-carbon (Cα) atom RMSDs of the third helices of R2 and R3 (R2R3) were calculated for complex of R2R3 binding with Moebius syndrome 1 (MBS1) at 300 K and complexes F and R at 300 K. (b) Complex of R2R3 binding with MBS1 and complexes F and R at 330 K. Complexes F and R for two pairs of ensembles are marked with black and red lines, respectively. The complex of R2R3 binding with MBS1 is represented by a green line.

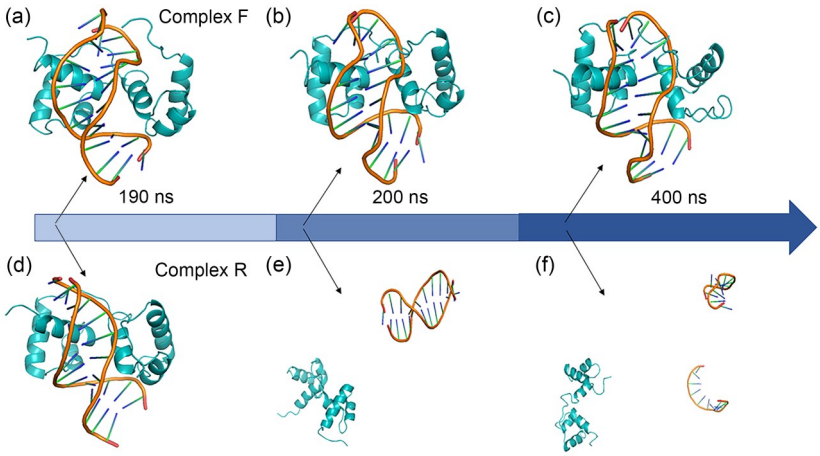

Fig. 3 Conformations of complexes F and R at 330 K. Conformations of complexes F and R were extracted at 190 ns (a, d), 200 ns (b, e), and 400 ns (c, f).

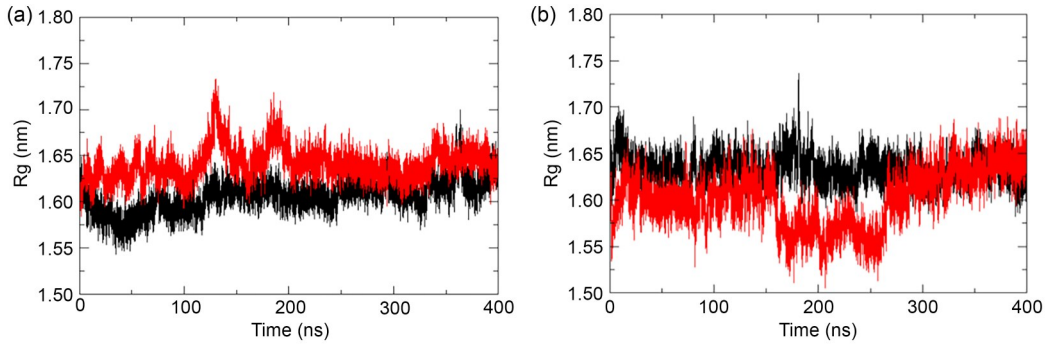

Fig. 4 Gyration radius of complexes F and R at 300 K (a) and 330 K (b). Complexes F and R for two pairs of ensembles are marked with black and red lines, respectively. Rg: radius of gyration.

300 K (Fig. S1a). In contrast, the three regions of RMSF between complexes F and R were significantly different at 330 K, corresponding to the first helix of R2 (Asp100-Try110), the third helix of R2 (Lys128-Trp147), and the third helix of R3 (Leu173-Asn183), respectively (Fig. S1b). This indicated that the conformational rearrangement took place in complex R at 330 K, and the most prominent change

occurred at the third helices of R2R3. Previous studies have suggested that the flexibilities of R2 and R3 helices are crucial for DNA binding (Ogata et al., 1992, 1994, 1995, 1996; Kanei-Ishii et al., 1994; Madan et al., 1995).

3.2 Principal component analyses for complexes F and R

In order to assess the potential difference of the motions between complexes F and R, the prominent characteristic motions during MD simulation were analyzed using PCA. The percentages of the top two eigenvectors of the motions in complexes F and R at 300 and 330 K were determined as 62%, 70%, 60%, and 59%, respectively. Our observation focused on the two recognition helices of R2 and R3. Notably, the flexible motion of the recognition helix of R3 in complex R at 330 K was quite different from that in other three complexes (Fig. 5). Specifically, the role of two recognition helices of R2 and R3 appeared like a clamp to hold the DNA. We noticed that the recognition helix of R3 was dissociated from the recognition helix of R2 in complex R at 330 K, which resulted in R2R3 losing the clamp function. A similar motion was observed in the P53 DNA complex (Pan and Nussinov, 2007; Dai and Yu, 2020). To validate our observation in PCA, we further analyzed the distance of the recognition helices between R2 and R3 in complexes F and R at 300 and 330 K. The distance of the two recognition helices in complex R at 330 K underwent a remarkable increase between 135 and 280 ns, compared to the distance of the recognition helices in complex F at 330 K (Fig. S2). Interestingly, the distance variation of the recognition-helices in complex R at 330 K matched that of the gap of gyration-radiuses between complexes F and R at 330 K. Thus, the opening motion of recognition-helices was the primary conformational variation in complex R at 330 K, which is the direct reason for the dissociation of c-Myb from DNA. However, it remained unclear which fundamental element triggered the opening motion of two recognition helices in complex R at 330 K.

3.3 Surface charge distribution

To explore the fundamental element that triggered the dissociation of c-Myb from DNA in complex R at 330 K, we analyzed the electrostatic potential on the surfaces of complexes F and R. As shown in Fig. 6, the recognition helices of R2R3 comprised the main region of the electrostatic potential distribution that acted as a positive charge pocket, which was the most plausible binding site for the negatively charged DNA. Interestingly, the differences in electrostatic potential in complex F increased from 72.826 to 80.552 continuously during the whole simulation process. In contrast, the differences in electrostatic potential in complex R decreased from 73.096 to 69.520 gradually during the first 190 ns before the dissociation of DNA from R2R3. However, after 190 ns, the

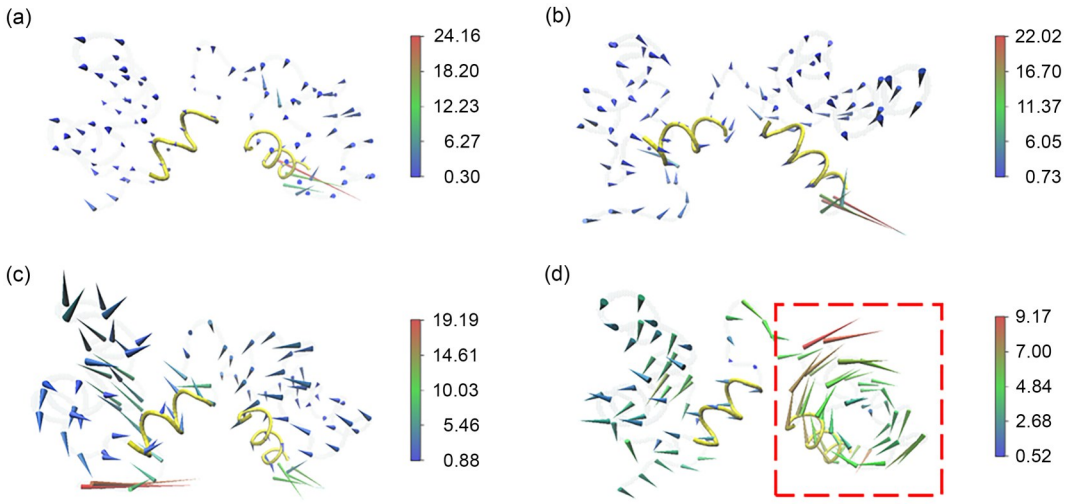

Fig. 5 Dominant motions in four simulation ensembles using principal component analysis. The porcupine plot of the top two eigenvectors in complex F at 300 K (a), complex R at 300 K (b), complex F at 330 K (c), and complex R at 330 K (d). The recognition helices of R2 and R3 are marked in yellow.

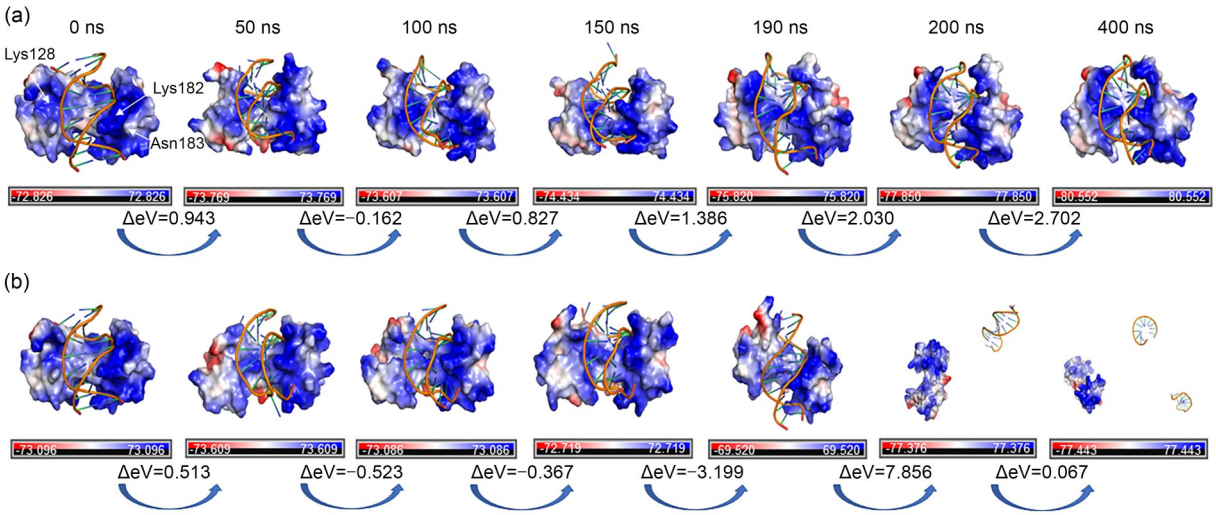

Fig. 6 Electrostatic potential differences (ΔeV) of complexes F and R in the whole molecular dynamics (MD) simulation process. (a) Complex F at 330 K in the whole MD simulation process; (b) Complex R at 330 K in the whole MD simulation process.

electrostatic potential differences quickly raised from 69.520 to 77.376. Notably, there was a dramatic reduction of 3.199 in the differences in electrostatic potential between 150 to 190 ns. The above results indicated that the variation in the electrostatic potential differences affected the interaction of c-Myb and DNA. A similar conclusion was drawn in the interaction of human RNA-binding motif single stranded interacting protein 1 (RBMS1) and *c-myc* proto-oncogene, namely the recognition of specific target sequence requires a coordination between conformational rearrangement and thermodynamics (Aggarwal and Bhavesh, 2021). Furthermore, we speculated that a reduced electrostatic potential on the surface of R2R3 can trigger the dissociation of DNA from c-Myb.

3.4 DNA conformation analysis

In order to investigate effect of the conformation of DNA on the interaction of c-Myb and *mim-1*, we performed DNA conformational analysis using Curves+ for four simulation systems of complexes F and R at 300 and 330 K. Specifically, the widths of the major and minor grooves of DNA were measured based on each 10-ns structure during the plateau of MD simulations, which corresponds to 100 to 400 ns for the 300-K systems, and 30 to 80 ns for the 330-K systems. As shown in Fig. 7, the widths of major and minor grooves in complex R at both 300 and 330 K were larger than those in complex F. Interestingly, the average width of the major groove in complex R at 330 K was significantly larger than that in complex R at 300 K, which was the same for the complex F at 300 and 330 K. The recent studies showed that the increase of the DNA groove width affected the interaction of Watson-Crick base pairs in DNA, which resulted in a decreased melting point of DNA (Padroni et al., 2019; Sarkar and Singh, 2020). Thus, we speculated that the increased major groove width of DNA in complex R at 330 K affected its capability to form hydrogen bonds between R2R3 and DNA.

3.5 Hydrogen-bond interaction of *mim-1* binding to c-Myb

With the aim to verify our hypothesis, the hydrogen-bond interactions of R2R3 and DNA were measured for complexes F and R at both 300 and 330 K at different time points (Table S4). The hydrogen bonds involved in Lys128, Lys182, and Asn183 play an essential role in the interaction of R2R3 and DNA (Ogata et al., 1994). As expected, the key hydrogen bonds in complex R at 330 K decreased gradually and disappeared at 190 ns before DNA dissociated from c-Myb, while the key hydrogen bonds were still kept in the other three complexes. The above results confirmed our hypothesis that the increased major groove width of DNA in complex R at 330 K affected the ability of hydrogen-bond formation between R2R3 and DNA, which resulted in the dissociation of DNA from R2R3.

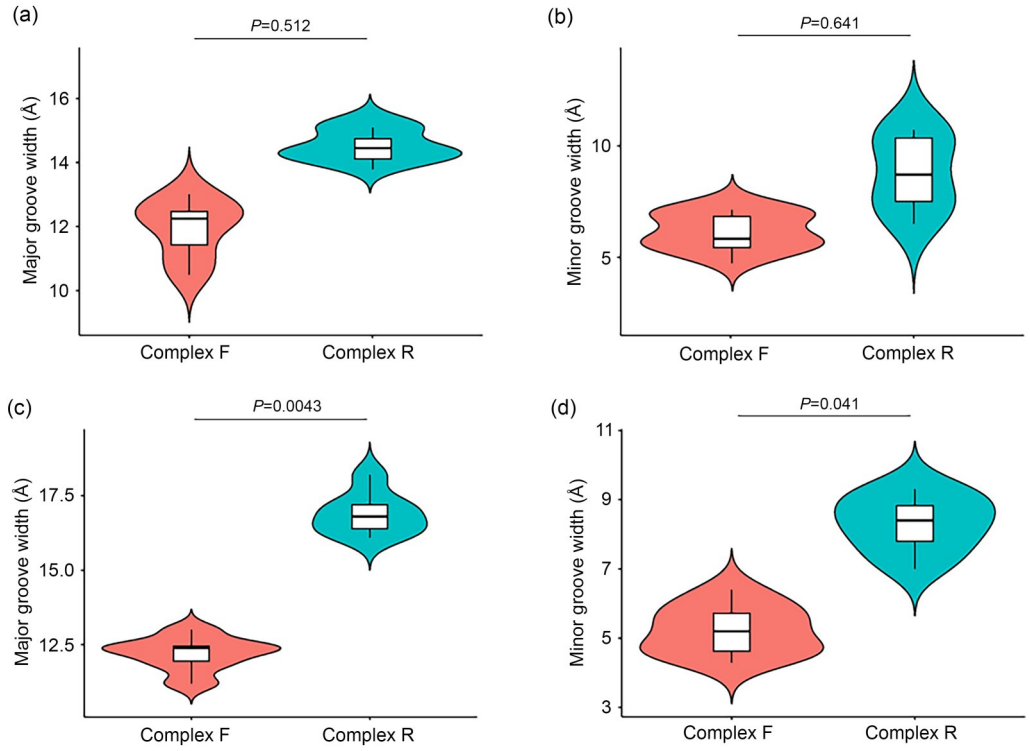

Fig. 7 DNA groove width of complexes F and R in plateau of molecular dynamics (MD) simulations. (a) Major groove width of complexes F and R at 300 K; (b) Minor groove width of complexes F and R at 300 K; (c) Major groove width of complexes F and R at 330 K; (d) Minor groove width of complexes F and R at 330 K. *P* value was calculated using unpaired *t*-test method, which evaluates the difference of the average groove widths between complexes F and R during plateau of MD simulations. 1 Å=1×10^{-10} m.

3.6 SMD simulations of complexes F and R

Our MD simulation results suggested that the electrostatic potential, major groove width, and hydrogen bonds made major contribution to DNA-specific recognition. To validate this proposition, SMD simulation studies were performed on complexes F and R. Specifically, the c-Myb protein served as the immobile reference, and we compared the DNA pulling forces between complexes F and R in 310 K during the 300-ps SMD simulation processes. As a result, the harmonic force used for complex F was 2150.20 kcal/(mol·nm), which was larger than that (1925.53 kcal/(mol·nm)) used for complex R (Fig. 8a). The time point for the dissociation of DNA from R2R3 occurred at 150 ps for complex F and at 214 ps for complex R (Fig. 8). As expected, the electrostatic potential surfaces of R2R3 in both complexes were decreased dramatically in both complexes before the time point of DNA dissociation (Fig. S3). The major groove width of both complexes increased gradually before the time point of DNA dissociation and dropped after the time point of DNA dissociation (Fig. S4). In addition, the key hydrogen bonds disappeared before the time point of DNA dissociation in both complexes (Table S5). Thus, the above results confirmed our conclusion from MD simulation studies.

3.7 Binding specificity of c-Myb to *mim-1* in vitro

In order to validate the MD simulation results and identify the binding specificity of c-Myb to the pseudo-palindromic consensus of *mim-1*, the binding titration analyses of c-Myb to *mim-1* and MBS1 were performed individually. The binding kinetic patterns (Fig. 9) showed that both reactions possessed similar binding titration curves and K_D values, indicating that both reactions not only had the same binding sites, but also seemed to be the single-binding site recognized by R2R3. To further determine whether cooperation occurred between the dual-Myb-binding sites of *mim-1*, the two titration curves were subjected to Hill plot analyses. The linear pattern of the regressive curves and the Hill coefficients both implied that no

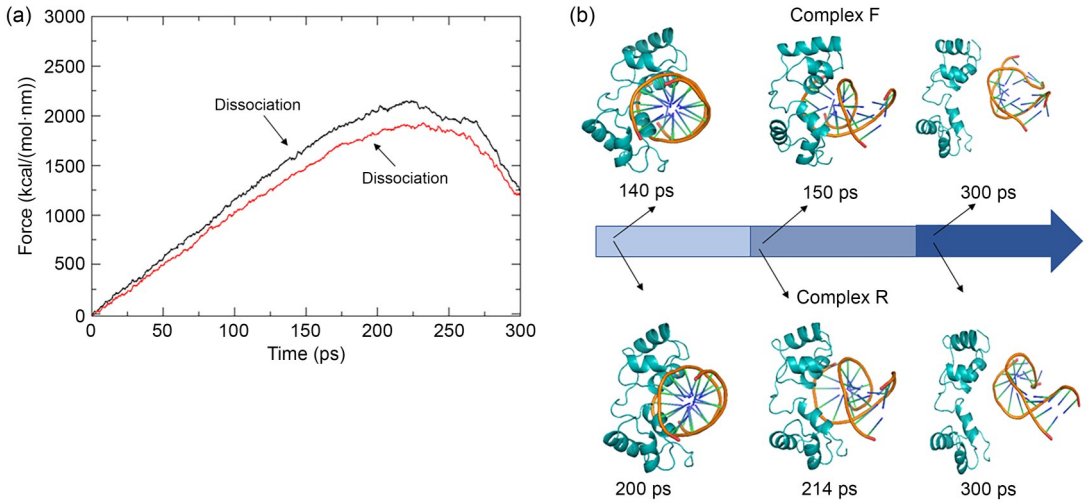

Fig. 8 Steered molecular dynamics (SMD) simulations of complexes F and R. (a) The pulling forces analyzed for complexes F and R which are marked with black and red lines, respectively. (b) The conformations of complexes F and R at different time points in SMD simulation studies.

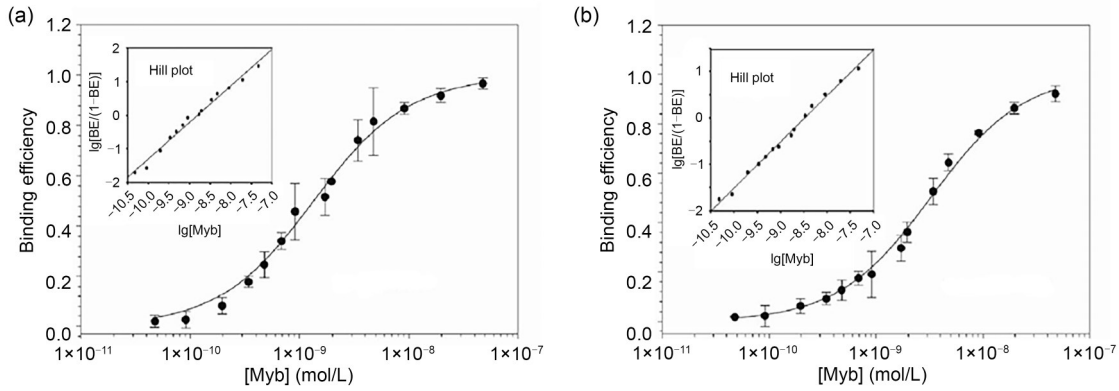

Fig. 9 Binding titration and Hill plot analyses of R2R3 binding to mitochondrial import protein 1 (*mim-1*) (a) or Moebius syndrome 1 (MBS1) (b). The K_D values were determined by quantitative electrophoretic mobility shift assay (EMSA). Data are expressed as mean±standard deviation (SD), n=3. BE: binding efficiency; [Myb]: concentration of Myb.

cooperative binding occurred in the reaction of R2R3 and the binding sites of *mim-1*. Thus, the above results indicated that only one of dual-Myb-binding sites in *mim-1* was recognized by R2R3, which was consistent with the MD simulation results.

In order to further validate whether c-Myb specifically binds to the forward strand between the dual-Myb-binding sites of *mim-1*, the variation in the binding free energy $\Delta\Delta G$ as a result of the single-base substitution was calculated based on the equilibrium dissociation constant K_D measured by quantitative EMSA. Specifically, each base pair of the *mim-1* DNA fragment from A4 to T10 was substituted by the other three kinds of corresponding base pairs (Fig. 10). The pattern of binding free energy $\Delta\Delta G$ showed that the mutants with the single-base substitution at A4, A5, and C6 caused obvious increases in the binding free energy (around 4 kcal/mol, corresponding to a ⩾1000-fold variation in binding affinity), indicating that the AAC positions acted as the most conserved core of the 5'-AACGG-3' consensus and provided highly specific interaction with R2R3. However, the single-base substitutions at other positions (from G7 to T10) resulted in relatively low $\Delta\Delta G$ values, except for the base G8 with a 2-fold $\Delta\Delta G$ increase, which seemed to be of moderate importance to the binding specificity. Notably, the values of $\Delta\Delta G$ at the positions (G8, T9, and T10) corresponding to the core of 5'-AACCG-3' were relatively low, implying that the complementary strand of the dual-Myb-binding sites was unlikely to

play an important role in the Myb-binding affinity. Thus, the above results demonstrated that c-Myb exclusively bound to the forward strand of the dual-Myb-binding sites, which further confirmed our MD simulation conclusion.

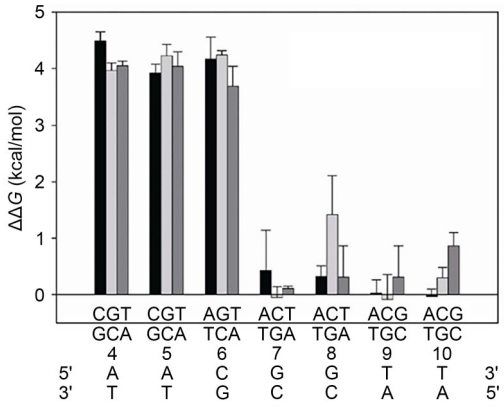

Fig. 10 Relative binding free energy changes (ΔΔ*G*) in the binding of R2R3 to mitochondrial import protein 1 (*mim-1*) upon systematic base substitution in the pseudo-palindromic consensus. The pseudo-palindromic consensus sequences in *mim-1* are shown at the bottom. Each solid bar indicates the extent of ΔΔ*G* caused by the corresponding base substitution. The values of ΔΔ*G* were calculated from K_D values derived from quantitative electrophoretic mobility shift assay (EMSA). Data are expressed as mean±standard deviation (SD), *n*=3. K_D: equilibrium dissociation constant.

4 Conclusions

In this work, the chicken myeloid gene *mim-1* was chosen as a binding target to study the binding specificity between dual-Myb-binding sites. Six MD simulation studies in different biological scenarios and totaling 2400 ns in time-scale revealed that the binding of c-Myb with *mim-1* in the forward strand was more stable than that in the reverse strand. The PCA dynamics trajectory analyses and the electrostatic potential surface analyses suggested that the reduced electrostatic potential surface of R2R3 triggered an opening motion of the recognition helices of R2R3, resulting in the dissociation of DNA from c-Myb in complex R at 330 K. In addition, the DNA confirmation analyses suggested that the increased major groove width of DNA in complex R at 330 K affected the capability of hydrogen-bond formation between R2R3 and the DNA reverse strand, which in turn directly resulted in the dissociation of DNA from R2R3. Hence, the electrostatic potential, major groove width, and hydrogen-bond interaction all played essential roles in the DNA-specific recognition by c-Myb, which was further confirmed by SMD simulation studies. In vitro, the kinetic pattern and quantitative EMSA analyses validated the conclusion of simulation studies, that is, c-Myb specifically bound to *mim-1* in the forward strand. This study addressed the DNA-specific recognition mechanism by c-Myb in the structure level, which is beneficial for investigating the cell early differentiation and proliferation regulated by c-Myb and the prediction of novel c-Myb-binding motifs in tumorigenesis.

Data availability statement

The structure of c-Myb binding to DNA of MBS1 is available on the Protein Data Bank (PDB code: No. 1MSE).

Acknowledgments

This work was supported by the National Key Research and Development Program of China (Nos. 2022YFC2402900 and 2022YFC2402901), the Fundamental Research Funds for the Central Universities (No. 226-2022-00213), and the Joint Funds of the Zhejiang Provincial Natural Science Foundation of China (No. LHDMD23H300001). We thank Tsukuba Life Sciences Center, the Institute of Physical and Chemical Research (RIKEN), Japan for the assistance on the quantitative EMSA experiments accomplished by Prof. Dongyun HAO during his post-doctoral study. We also thank Dr. Shichen WANG from the Texas A&M AgriLife Research (TX, USA) for the assistance on bioinformatics experiments.

Author contributions

Shan WANG conceived, designed, and supervised the whole research, designed bioinformatics experiments, and wrote the manuscript with input from all authors. Dongyun HAO conceived and supervised the research, designed validation experiments, and wrote the manuscript. Jinru WENG carried out the bioinformatics analysis. Shuo YANG carried out the validation experiments. Jinkang SHEN, Hongsen LIU, and Yuzi XU participated in the bioinformatics analysis and interpretation of data. All authors have read and approved the final manuscript, and therefore, have full access to all the data in the study and take responsibility for the integrity and security of the data.

Compliance with ethics guidelines

Shan WANG is a young scientist committee member for *Journal of Zhejiang University-SCIENCE B (Biomedicine & Biotechnology)* and was not involved in the editorial review or the decision to publish this article. Jinru WENG, Shuo YANG, Jinkang SHEN, Hongsen LIU, Yuzi XU, Dongyun HAO, and Shan WANG declare that they have no conflicts of interest.

This article does not contain any studies with human or animal subjects performed by any of the authors.

References

Abaza MSI, Al-Attiyah RJ, Al-Saffar AM, et al., 2004. Antisense oligodeoxynucleotide directed against c-myb has anticancer activity and potentiates the antiproliferative effect of conventional anticancer drugs acting by different mechanisms in human colorectal cancer cells. *Tumor Biol*, 24(5):241-257. https://doi.org/10.1159/000076139

Aggarwal P, Bhavesh NS, 2021. Hinge like domain motion facilitates human RBMS1 protein binding to proto-oncogene *c-myc* promoter. *Nucleic Acids Res*, 49(10):5943-5955. https://doi.org/10.1093/nar/gkab363

Berendsen HJC, van der Spoel D, van Drunen R, 1995. GROMACS: a message-passing parallel molecular dynamics implementation. *Comp Phys Comm*, 91(1-3):43-56. https://doi.org/10.1016/0010-4655(95)00042-E

Bernstein FC, Koetzle TF, Williams GJB, et al., 1978. The protein data bank: a computer-based archival file for macromolecular structures. *Arch Biochem Biophys*, 185(2):584-591. https://doi.org/10.1016/0003-9861(78)90204-7

Bhattarai G, Lee YH, Lee MH, et al., 2013. Gene delivery of c-myb increases bone formation surrounding oral implants. *J Dent Res*, 92(9):840-845. https://doi.org/10.1177/0022034513497753

Biedenkapp H, Borgmeyer U, Sippel AE, et al., 1988. Viral *myb* oncogene encodes a sequence-specific DNA-binding activity. *Nature*, 335(6193):835-837.

Cicirò Y, Sala A, 2021. MYB oncoproteins: emerging players and potential therapeutic targets in human cancer. *Oncogenesis*, 10(2):19. https://doi.org/10.1038/s41389-021-00309-y

Clesham K, Walf-Vorderwülbecke V, Gasparoli L, et al., 2022. Identification of a c-MYB-directed therapeutic for acute myeloid leukemia. *Leukemia*, 36(6):1541-1549. https://doi.org/10.1038/s41375-022-01554-9

Dai LQ, Yu J, 2020. Inchworm stepping of Myc-Max heterodimer protein diffusion along DNA. *Biochem Biophys Res Commun*, 533(1):97-103. https://doi.org/10.1016/j.bbrc.2020.08.004

Essmann U, Perera L, Berkowitz ML, et al., 1995. A smooth particle mesh Ewald method. *J Chem Phys*, 103(19):8577-8593. https://doi.org/10.1063/1.470117

Fuglerud BM, Lemma RB, Wanichawan P, et al., 2017. A c-Myb mutant causes deregulated differentiation due to impaired histone binding and abrogated pioneer factor function. *Nucleic Acids Res*, 45(13):7681-7696. https://doi.org/10.1093/nar/gkx364

Gautam S, Fioravanti J, Zhu W, et al., 2019. The transcription factor c-Myb regulates $CD8^+$ T cell stemness and antitumor immunity. *Nat Immunol*, 20(3):337-349. https://doi.org/10.1038/s41590-018-0311-z

Gonda TJ, Leo P, Ramsay RG, 2008. Estrogen and *MYB* in breast cancer: potential for new therapies. *Expert Opin Biol Ther*, 8(6):713-717. https://doi.org/10.1517/14712598.8.6.713

Graf T, 1992. Myb: a transcriptional activator linking proliferation and differentiation in hematopoietic cells. *Curr Opin Genet Dev*, 2(2):249-255. https://doi.org/10.1016/S0959-437X(05)80281-3

Greig KT, de Graaf CA, Murphy JM, et al., 2010. Critical roles for c-Myb in lymphoid priming and early B-cell development. *Blood*, 115(14):2796-2805. https://doi.org/10.1182/blood-2009-08-239210

Hao DY, Ohme-Takagi M, Sarai A, 1998. Unique mode of GCC box recognition by the DNA-binding domain of ethylene-responsive element-binding factor (ERF domain) in plant. *J Biol Chem*, 273(41):26857-26861. https://doi.org/10.1074/jbc.273.41.26857

Hao DY, Yamasaki K, Sarai A, et al., 2002. Determinants in the sequence specific binding of two plant transcription factors, CBF1 and NtERF2, to the DRE and GCC motifs. *Biochemistry*, 41(13):4202-4208. https://doi.org/10.1021/bi015979v

Hess B, Bekker H, Berendsen HJC, et al., 1997. LINCS: a linear constraint solver for molecular simulations. *J Comput Chem*, 18(12):1463-1472. https://doi.org/10.1002/(SICI)1096-987X(199709)18:12<1463::AID-JCC4>3.0.CO;2-H

Heuser C, Gattinoni L, 2022. c-Myb redefines the hierarchy of stem-like T cells. *Nat Immunol*, 23(10):1405-1407. https://doi.org/10.1038/s41590-022-01319-7

Hugo H, Cures A, Suraweera N, et al., 2006. Mutations in the *MYB* intron I regulatory sequence increase transcription in colon cancers. *Genes Chromosomes Cancer*, 45(12):1143-1154. https://doi.org/10.1002/gcc.20378

Kanei-Ishii C, Yasukawa T, Morimoto RI, et al., 1994. c-Myb-induced *trans*-activation mediated by heat shock elements without sequence-specific DNA binding of c-Myb. *J Biol Chem*, 269(22):15768-15775. https://doi.org/10.1016/S0021-9258(17)40747-2

Katzen AL, Kornberg TB, Bishop JM, 1985. Isolation of the proto-oncogene *c-myb* from D. melanogaster. *Cell*, 41(2):449-456. https://doi.org/10.1016/S0092-8674(85)80018-0

Kawasaki M, Oda M, 2021. DNA-binding function of c-Myb R2R3 around thermal denaturation temperature. *Biophys Physicobiol*, 18:78-84. https://doi.org/10.2142/biophysico.bppb-v18.009

Krissinel E, Henrick K, 2007. Inference of macromolecular assemblies from crystalline state. *J Mol Biol*, 372(3):774-797. https://doi.org/10.1016/j.jmb.2007.05.022

Lauria A, Ippolito M, Almerico AM, 2009. Principal component analysis on molecular descriptors as an alternative point of view in the search of new Hsp90 inhibitors. *Comput Biol Chem*, 33(5):386-390. https://doi.org/10.1016/j.compbiolchem.2009.07.010

Lavery R, Moakher M, Maddocks JH, et al., 2009. Conformational analysis of nucleic acids revisited: Curves+. *Nucleic Acids Res*, 37(17):5917-5929. https://doi.org/10.1093/nar/gkp608

Lee YH, Kim HS, Kim JS, et al., 2016. C-myb regulates autophagy for pulp vitality in glucose oxidative stress. *J Dent Res*, 95(4):430-438.

https://doi.org/10.1177/0022034515622139

Lindorff-Larsen K, Piana S, Palmo K, et al., 2010. Improved side-chain torsion potentials for the Amber ff99SB protein force field. *Proteins*, 78(8):1950-1958. https://doi.org/10.1002/prot.22711

Lu H, Wang Y, Huang Y, et al., 2013. Expression and prognostic role of c-Myb as a novel cell cycle protein in esophageal squamous cell carcinoma. *Clin Transl Oncol*, 15(10):796-801. https://doi.org/10.1007/s12094-013-1009-1

Madan A, Radha PK, Hosur RV, et al., 1995. Bacterial expression, characterization and DNA binding studies on *Drosophila melanogaster* c-Myb DNA-binding protein. *Eur J Biochem*, 232(1):150-158. https://doi.org/10.1111/j.1432-1033.1995.tb20793.x

Malaterre J, Mantamadiotis T, Dworkin S, et al., 2008. c-Myb is required for neural progenitor cell proliferation and maintenance of the neural stem cell niche in adult brain. *Stem Cells*, 26(1):173-181. https://doi.org/10.1634/stemcells.2007-0293

Mansour MR, Abraham BJ, Anders L, et al., 2014. Oncogene regulation. An oncogenic super-enhancer formed through somatic mutation of a noncoding intergenic element. *Science*, 346(6215):1373-1377. https://doi.org/10.1126/science.1259037

Maurice D, Hooper J, Lang G, et al., 2007. c-myb regulates lineage choice in developing thymocytes via its target gene Gata3. *EMBO J*, 26(15):3629-3640. https://doi.org/10.1038/sj.emboj.7601801

Mitani Y, Li J, Rao PH, et al., 2010. Comprehensive analysis of the *MYB-NFIB* gene fusion in salivary adenoid cystic carcinoma: incidence, variability, and clinicopathologic significance. *Clin Cancer Res*, 16(19):4722-4731. https://doi.org/10.1158/1078-0432.CCR-10-0463

Miyazaki T, Pan Y, Joshi K, et al., 2012. Telomestatin impairs glioma stem cell survival and growth through the disruption of telomeric G-quadruplex and inhibition of the proto-oncogene, *c-Myb*. *Clin Cancer Res*, 18(5):1268-1280. https://doi.org/10.1158/1078-0432.CCR-11-1795

Morris GM, Huey R, Lindstrom W, et al., 2009. AutoDock4 and AutoDockTools4: automated docking with selective receptor flexibility. *J Comput Chem*, 30(16):2785-2791. https://doi.org/10.1002/jcc.21256

Ness SA, 1996. The Myb oncoprotein: regulating a regulator. *Biochim Biophys Acta*, 1288(3):F123-F139. https://doi.org/10.1016/s0304-419x(96)00027-3

Ness SA, Marknell Å, Graf T, 1989. The v-*myb* oncogene product binds to and activates the promyelocyte-specific *mim-1* gene. *Cell*, 59(6):1115-1125. https://doi.org/10.1016/0092-8674(89)90767-8

Ni SJ, Zhu JY, Zhang JG, et al., 2015. Expression and clinical role of NF45 as a novel cell cycle protein in esophageal squamous cell carcinoma (ESCC). *Tumor Biol*, 36(2):747-756. https://doi.org/10.1007/s13277-014-2683-5

Nishina Y, Nakagoshi H, Imamoto F, et al., 1989. *Trans*-activation by the c-*myb* proto-oncogene. *Nucleic Acids Res*, 17(1):107-117. https://doi.org/10.1093/nar/17.1.107

O'Boyle NM, Banck M, James CA, et al., 2011. Open Babel: an open chemical toolbox. *J Cheminform*, 3:33. https://doi.org/10.1186/1758-2946-3-33

Ogata K, Hojo H, Aimoto S, et al., 1992. Solution structure of a DNA-binding unit of Myb: a helix-turn-helix-related motif with conserved tryptophans forming a hydrophobic core. *Proc Natl Acad Sci USA*, 89(14):6428-6432. https://doi.org/10.1073/pnas.89.14.6428

Ogata K, Morikawa S, Nakamura H, et al., 1994. Solution structure of a specific DNA complex of the Myb DNA-binding domain with cooperative recognition helices. *Cell*, 79(4):639-648. https://doi.org/10.1016/0092-8674(94)90549-5

Ogata K, Morikawa S, Nakamura H, et al., 1995. Comparison of the free and DNA-complexed forms of the DMA-binding domain from c-Myb. *Nat Struct Biol*, 2(4):309-320. https://doi.org/10.1038/nsb0495-309

Ogata K, Kanei-Ishii C, Sasaki M, et al., 1996. The cavity in the hydrophobic core of Myb DNA-binding domain is reserved for DNA recognition and *trans*-activation. *Nat Struct Biol*, 3(2):178-187. https://doi.org/10.1038/nsb0296-178

Oh IH, Reddy EP, 1999. The *myb* gene family in cell growth, differentiation and apoptosis. *Oncogene*, 18(19):3017-3033. https://doi.org/10.1038/sj.onc.1202839

Ording E, Kvåvik W, Bostad A, et al., 1994. Two functionally distinct half sites in the DNA-recognition sequence of the Myb oncoprotein. *Eur J Biochem*, 222(1):113-120. https://doi.org/10.1111/j.1432-1033.1994.tb18848.x

Padroni G, Withers JM, Taladriz-Sender A, et al., 2019. Sequence-selective minor groove recognition of a DNA duplex containing synthetic genetic components. *J Am Chem Soc*, 141(24):9555-9563. https://doi.org/10.1021/jacs.8b12444

Pan JH, Adair-Kirk TL, Patel AC, et al., 2014. Myb permits multilineage airway epithelial cell differentiation. *Stem Cells*, 32(12):3245-3256. https://doi.org/10.1002/stem.1814

Pan YP, Nussinov R, 2007. Structural basis for p53 binding-induced DNA bending. *J Biol Chem*, 282(1):691-699. https://doi.org/10.1074/jbc.M605908200

Pardo-Saganta A, Law BM, Tata PR, et al., 2015. Injury induces direct lineage segregation of functionally distinct airway basal stem/progenitor cell subpopulations. *Cell Stem Cell*, 16(2):184-197. https://doi.org/10.1016/j.stem.2015.01.002

Patt S, Thiel G, Maas S, et al., 1993. Chromosomal changes and correspondingly altered proto-oncogene expression in human gliomas. Value of combined cytogenetic and molecular genetic analysis. *Anticancer Res*, 13(1):113-118.

Paz-Ares J, Ghosal D, Wienand U, et al., 1987. The regulatory c1 locus of Zea mays encodes a protein with homology to myb proto-oncogene products and with structural similarities to transcriptional activators. *EMBO J*, 6(12):3553-3558. https://doi.org/10.1002/j.1460-2075.1987.tb02684.x

Ramsay RG, Gonda TJ, 2008. MYB function in normal and cancer cells. *Nat Rev Cancer*, 8(7):523-534. https://doi.org/10.1038/nrc2439

Rohs R, West SM, Sosinsky A, et al., 2009. The role of DNA shape in protein-DNA recognition. *Nature*, 461(7268):1248-1253.

https://doi.org/10.1038/nature08473

Rosson D, Reddy EP, 1986. Nucleotide sequence of chicken *c-myb* complementary DNA and implications for *myb* oncogene activation. *Nature*, 319(6054):604-606.

Sarkar S, Singh PC, 2020. Alteration of the groove width of DNA induced by the multimodal hydrogen bonding of denaturants with DNA bases in its grooves affects their stability. *Biochim Biophys Acta Gen Subj*, 1864(3):129498. https://doi.org/10.1016/j.bbagen.2019.129498

Tan FE, Vladar EK, Ma L, et al., 2013. *Myb* promotes centriole amplification and later steps of the multiciliogenesis program. *Development*, 140(20):4277-4286. https://doi.org/10.1242/dev.094102

Tanikawa J, Yasukawa T, Enari M, et al., 1993. Recognition of specific DNA sequences by the c-*myb* protooncogene product: role of three repeat units in the DNA-binding domain. *Proc Natl Acad Sci USA*, 90(20):9320-9324. https://doi.org/10.1073/pnas.90.20.9320

Teleman O, Jönsson B, Engström S, 1987. A molecular dynamics simulation of a water model with intramolecular degrees of freedom. *Mol Phys*, 60(1):193-203. https://doi.org/10.1080/00268978700100141

Torelli G, Venturelli D, Coló A, et al., 1987. Expression of c-*myb* protooncogene and other cell cycle-related genes in normal and neoplastic human colonic mucosa. *Cancer Res*, 47(20):5266-5269.

Wallrapp C, Müller-Pillasch F, Solinas-Toldo S, et al., 1997. Characterization of a high copy number amplification at 6q24 in pancreatic cancer identifies *c-myb* as a candidate oncogene. *Cancer Res*, 57(15):3135-3139.

Wang S, Yang S, An BY, et al., 2011. Molecular dynamics analysis reveals structural insights into mechanism of nicotine *N*-demethylation catalyzed by tobacco cytochrome P450 mono-oxygenase. *PLoS ONE*, 6(8):e23342. https://doi.org/10.1371/journal.pone.0023342

Wolff L, 1996. Myb-induced transformation. *Crit Rev Oncog*, 7(3-4):245-260. https://doi.org/10.1615/critrevoncog.v7.i3-4.60

Xu LH, Zhao F, Yang WW, et al., 2019. MYB promotes the growth and metastasis of salivary adenoid cystic carcinoma. *Int J Oncol*, 54(5):1579-1590. https://doi.org/10.3892/ijo.2019.4754

Yang LW, Eyal E, Bahar I, et al., 2009. Principal component analysis of native ensembles of biomolecular structures (PCA_NEST): insights into functional dynamics. *Bioinformatics*, 25(5):606-614. https://doi.org/10.1093/bioinformatics/btp023

Zhao L, Glazov EA, Pattabiraman DR, et al., 2011. Integrated genome-wide chromatin occupancy and expression analyses identify key myeloid pro-differentiation transcription factors repressed by Myb. *Nucleic Acids Res*, 39(11):4664-4679. https://doi.org/10.1093/nar/gkr024

Supplementary information

Tables S1–S5; Figs. S1–S4